AFRL-RI-RS-TR-2009-118
Final Technical Report
May 2009

MULTI-SENSOR VISION DATA FUSION FOR SMART AIRBORNE SURVEILLANCE

Tennessee State University

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88th ABW, Wright-Patterson AFB Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RI-RS-TR-2009-118 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR:

/s/ /s/

HENRY X. SIMMONS
Work Unit Manager

JAMES W. CUSACK, Chief
Information Systems Division
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
MAY 09	Final	Feb 08 – Dec 08

4. TITLE AND SUBTITLE

MULTI-SENSOR VISION DATA FUSION FOR SMART AIRBORNE SURVEILLANCE

5a. CONTRACT NUMBER
FA8750-08-1-0116

5b. GRANT NUMBER
N/A

5c. PROGRAM ELEMENT NUMBER
62702F

6. AUTHOR(S)

Aki Sekmen and Fenghui Yao

5d. PROJECT NUMBER
558B

5e. TASK NUMBER
HB

5f. WORK UNIT NUMBER
CU

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Tennessee State University
Department of Computer Science
3500 John A. Merritt Blvd.
Nashville, TN 37209

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFRL/RISA
525 Brooks Rd.
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)
N/A

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-RI-RS-TR-2009-118

12. DISTRIBUTION AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED. PA# 88ABW-2009-2029

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This research addresses the problem of detection and tracking of moving targets of interest within a multi-source data fusion framework that can elegantly integrate vision data captured by airborne optical and infrared (IR) cameras. The system can be employed in tactical airborne surveillance applications that are essential for activity analysis and situation awareness. Complementary information from the optical and IR cameras enables to perceive features in the environment more accurately and reliably. This report describes the research activities and developments during the course of the project.

15. SUBJECT TERMS
"Airborne Surveillance" "Vision Data fusion"

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Henry X. Simmons
U	U	U	UU	27	19b. TELEPHONE NUMBER (*Include area code*) N/A

Table of Contents

List of Figures

List of Tables

EXECUTIVE SUMMARY

This research addresses the problem of detection and tracking of moving targets of interest within a multi-source data fusion framework that can elegantly integrate vision data captured by airborne optical and infrared (IR) cameras. The system can be employed in tactical airborne surveillance applications that are essential for activity analysis and situation awareness. Complementary information from the optical and IR cameras enables to perceive features in the environment more accurately and reliably. This report describes the research activities and developments during the course of the project.

The airborne surveillance systems using complementary optical and IR cameras are well suited to surveillance over complex terrain. Image fusion enables certain features to be detected more accurately. Some features that are impossible to be perceived by any individual sensor may be distinguished. Optical cameras may have high dynamic range and higher resolution; however, they lack contrast between targets and background. Also, they fail in the presence of dust, fog, or smoke and require active illumination when light levels are low ambient. IR cameras exhibit a high contrast between the background and targets of interest. However, they have low resolution and they are not useful in the environments where the scene has a uniform temperature (such as the ground after rain).

In this report, a system that can combine optical and IR images generated from an airborne platform is described. The system also performs automatic target detection using the fused images. The objects within optical and IR images are first detected. Then, an object mapping to determine certain parameters for image fusion is performed. Finally, the optical and IR images are fused by utilizing Discrete Wavelet Transform (DWT) and the targets are detected using the fused image sequences. The real-world videos generated from an unmanned aerial vehicle (UAV) are used for system evaluation. The experiment results validate the proposed system.

Two (2) undergraduate students from the Department of Computer Science and two (2) graduate student from the Computer and Information Systems Engineering were partially supported by this project.

1. INTRODUCTION

This research developed a multi-source vision data fusion system for detection and tracking of moving targets from stationary and moving sensor platforms. Data from vision sensors (optical and IR cameras) were fused in data-level and feature-level for multi-modal data integration. The system can be employed in remote surveillance applications that are essential for activity analysis and situation awareness. A smart surveillance system is expected to detect, identify, and track possible targets of interest autonomously. This research addressed only target detection and tracking using optical and IR cameras together.

This report presents the description of a novel system that can integrate optical and IR images and then use the fused image sequences for moving target detection.

1.1. OVERVIEW

Image fusion is a process of combining multiple images to form a single image by utilizing certain features from each image. The successful fusion of images acquired from different modalities or instruments is of great importance in many applications such as image analysis and computer vision [1], [2], [3], concealed weapon detection [4], [5], and autonomous landing guidance [6], [7]. Image fusion can be performed at four levels of the information representation, which are signal, pixel, feature, and symbolic levels. Multi-scale transforms are widely used for analyzing the information content of images for image fusion. Several multiscale transforms have become very popular. These include the Laplacian pyramid transform [8], the contrast pyramid transform [9], the gradient pyramid transform [10], and the discrete wavelet transform (DWT) [11]. A comparative study of these methods is given in [12]. Recently, a new method that is based on trajectory association is proposed for image fusion [13]. Many of these works hand the still images. This paper describes a novel approach for fusing optical an`d infrared (IR) image sequences collected by an airborne platform and its application on target detection. A new algorithm is proposed for the effective fusion of airborne images from heterogeneous cameras. First, moving objects within the optical and IR images are detected. Second, an object mapping process is applied to map the objects in the optical images with the object in the IR images to find a relation between the images. Third, the optical and IR images are fused and finally moving targets are detected using the fused image sequences. The main contribution of this work is the development and evaluation of a novel algorithm for fusion of the airborne optical and IR images that results in more effective target detection. The foci of this algorithm are the object-based image fusion and target localization.

1.2. PARTICIPANTS

Two (2) faculty members, the PI and the Co-PI, were actively involved with the following contributions. An extensive search for undergraduate and graduate student participation as well as a post doctoral researcher or a research associate was pursued. Two (2) undergraduate graduate students from the College of Engineering, Computer Science and Technology were recruited. Two (2) graduate students were also partially supported to help developing some computer vision programs. One of the undergraduate students worked on his Senior Project, which was a direct product of this project.

1.3. PUBLICATIONS

The following conference paper is direct product of this research project.

> "Multiple-source airborne IR and optical image fusion and its application to target detection," F. Yao and A. Sekmen, *4th International Symposium on Visual Computing*, Las Vegas, NV, USA, December 2008.

In addition to this, Jeffrey Boyland, a Computer Science undergraduate student, has developed his Senior Project titled "Moving Target Detection in IR Sequences for Airborne Platforms".

1.4. OUTLINE OF THE REPORT

The report is organized as follows: Section 2 presents the novel image fusion and target detection system. The experimental results are described in Section 3. The performance analysis is given in Section 4 and some conclusions are given and future work is motivated in Section 5.

2. SYSTEM DESCRIPTION

The goal of the proposed research is to develop and evaluate a multi-source vision sensor data fusion framework for automatic detection and tracking of targets of interest for smart video surveillance. The research have implemented and evaluated multi-source sensor fusion algorithms for fast target detection and tracking that can operate in real-time or near real-time under combat environments.

This work discusses the fusion of images generated by an optical camera and an IR camera mounted on a helicopter, and the target detection from the fused images. Fig. 1 shows the samples of an optical image and an IR image. The fusion of these two types of images faces the following problems.

(a) (b)

Fig.1. (a) 640×480 color optical image; (b) 320×256 IR image.

(i) Everything in the scene including background appears to be moving since the cameras are mounted on a moving platform;

(ii) The optical image is color image, and IR image is grayscale image but recorded as pseudo color image, *i.e.*, IR signature is recorded to R-, G-, and B-channels. The resolution is different (640×480 for optical image and 320×256 for IR image), and the ratio of width to height is different;

(iii) There are some region overlaps, however, those regions are unknown;

(iv) There are multiple targets in images, and the number of targets may change (exit or reenter the field of view of a camera).

To address these problems, we designed an object mapping based image fusion and target detection algorithm. The entire processing flow is shown in Fig. 2. This algorithm consists of image registration, image fusion, and target detection. This research assumes that multiple cameras are mounted on the same helicopter or an unmanned aerial vehicle (UAV). Therefore, it is only necessary to perform the image registration once using certain number of image sequences to determine a relative motion relation between the optical and IR cameras. After performing the image registration, the registration parameters are used for image fusion and target detection. The following explains these three components in detail.

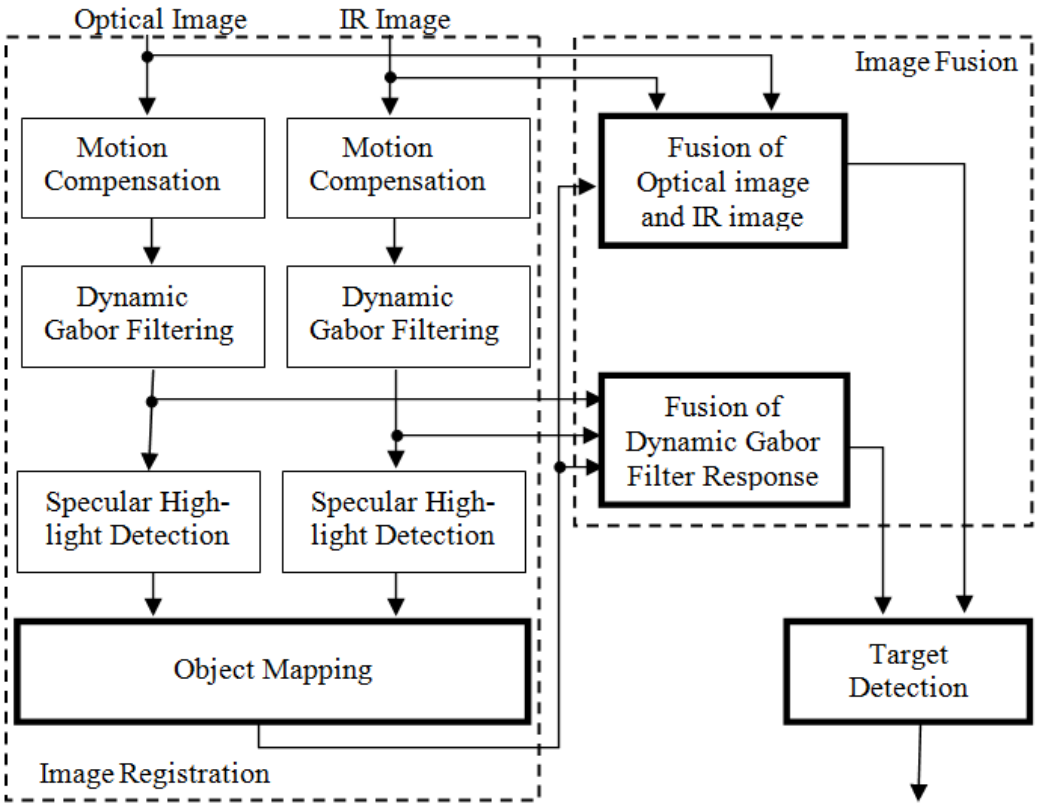


Fig.2. The process flow of the detection and fusion system.

2.1 IMAGE REGISTRATION

Image registration is the process of transforming the different sets of images into a common coordinate system. As shown in Fig. 2, the image registration in this system includes object detection from both optical and IR image, and object mapping. The object detection is based on the algorithm developed in our previous work [14]. This section first summarizes the object detection algorithm in Section 2.1.1. Then it mainly discusses the object mapping in Section 2.1.2.

2.1.1 Object Detection

The object detection contains motion compensation, dynamic Gabor filtering (DGF), and specular highlight detection. Let F_i^{τ} denote the *i*-th image frame, where $\tau \in \{O, I\}$, and *O* and *I* represents the optical image and IR image, respectively. Then the object detection algorithm following can be briefly summarized as follows. Details are referred to [14].

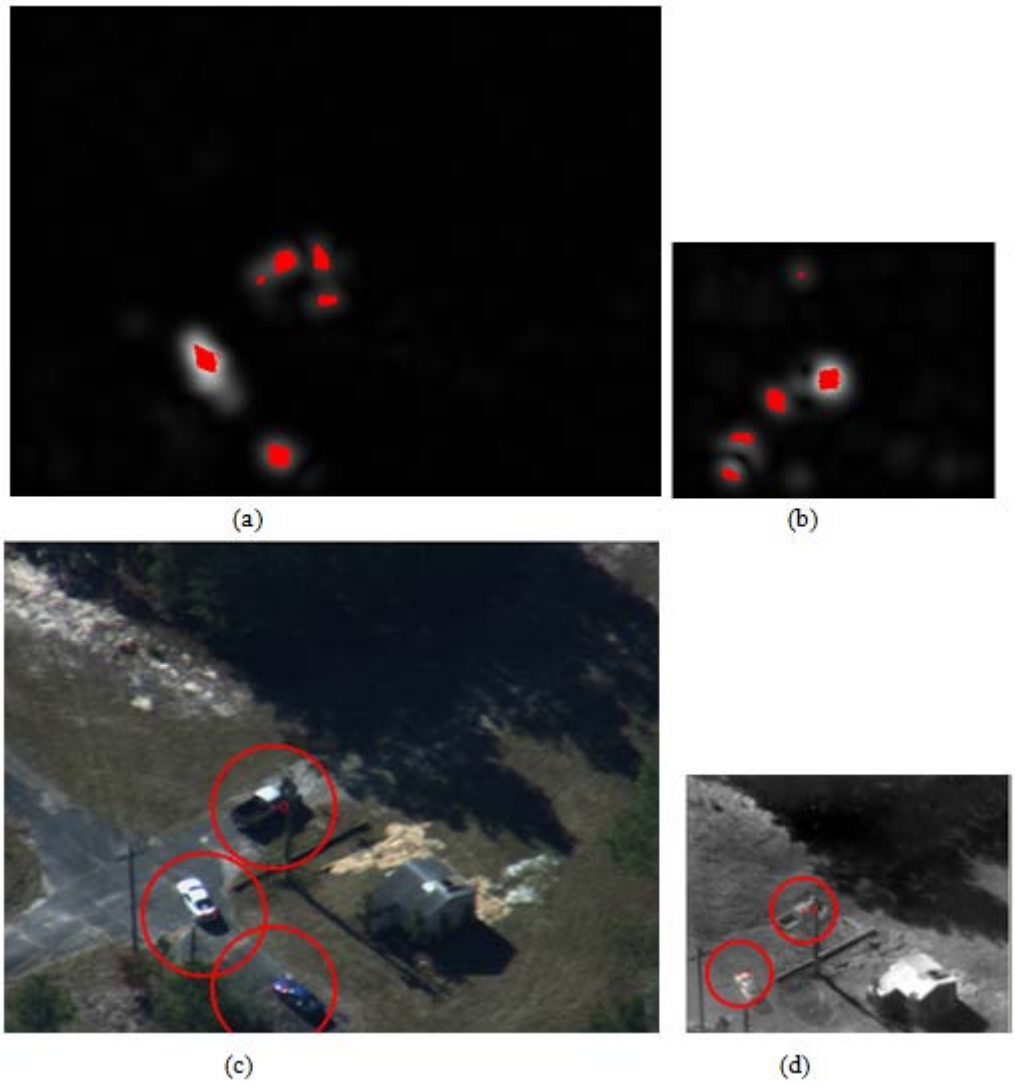


Fig.3. Specular highlight detection and object detection results. (a) and (b) Specular highlights detected from the optical image and IR image, respectively. (c) and (d) Objects detected from the optical image and IR image, respectively.

(i) For two consecutive frames, $F_{i-\Delta}^{\tau}$ and F_i^{τ} (Δ is the sampling interval), the feature points are detected by using Shi-Tomasi's method [15].

(ii) The optical flows between $F_{i-\Delta}^{\tau}$ and F_i^{τ} are detected by using Bouguet's algorithm [16]. The feature points are separated into inliers and outliers, where inliers are corresponding to the background, and outliers to the moving objects.

(iii) The inliers are used to estimate the affine transformation model between $F_{i-\Delta}^{\tau}$ and F_i^{τ} by using a RANSAC-like algorithm.

(iv) After the affine transformation model is determined, the frame difference is generated according to $F_{diff}^{\tau} = \left| F_i^{\tau} - \omega \times F_{i-\Delta}^{\tau} \right|$, where ω is the affine transformation model. Hence, the foreground can be separated from the background.

(v) DGF is applied to F_{diff}^{τ}, where the orientation of DGF is controlled by the optical flows corresponding to the inliers.

(vi) Specular highlight are detected. After DGF, the object detection becomes the detection of specular highlights. The detected highlights, after being filtered and merged, are considered as the objects.

The detected highlights and objects from the input images in Fig. 1 are shown in Fig. 3, where three objects are detected from the optical image, and two from IR image. In the following, the objects detected from the optical image and IR image are denoted by $O^O = \{O_1^O, O_2^O, ..., O_M^O\}$ and $O^I = \{O_1^I, O_2^I, ..., O_N^I\}$, respectively, where *M* and *N* are the number of objects in optical and IR images, respectively.

2.1.2 Object Mapping

As shown in Fig. 1, the optical image and IR image are different in resolution, size, and width-to-height ratio. The registration/fusion of these two images can be defined as,

$$F_{fuse} = F_i^O \oplus \tilde{F}_i^I(s, \theta, \boldsymbol{\lambda}), \tag{1}$$

where $\tilde{F}_i^I(s, \theta, \lambda)$ is the output image of the IR image F_i^I after being enlarged by scaling factor *s*, translated by the translation vector $\boldsymbol{\lambda}$, and rotated by angle θ, and $\oplus$ is the image fusion operator. The task of image registration is to find *s*, θ, and $\boldsymbol{\lambda}$, which will be discussed in this section. The task of image fusion is to find fusion operator $\oplus$, which will be discussed in Section 2.2.

To find s, θ, and $\boldsymbol{\lambda}$, we employ the brute force algorithm which is described below. For $O_m^O \in O^O$ and $O_n^I \in O^I$ (m = 1, 2, …, M, n = 1, 2, …, N), we extract the grayscale sub-image $I_{m,sub}^{O,L_O}$ from the color image F_i^O centered at C_m^O, and the grayscale sub-image $I_{n,sub}^{I,Li}$ from the pseudo color image F_i^I centered at C_n^I, respectively, where C_m^O is the center of the object O_m^O, and C_n^I of the object O^I, L_O is the size of $I_{m,sub}^{O,L_O}$, and L_I of $I_{n,sub}^{I,L_I}$. Note that L_I is smaller than L_O. Template matching for $I_{n,sub}^{I,Li}$ and $I_{m,sub}^{O,L_O}$ is performed in the following way. $I_{n,sub}^{I,Li}$ is shifted over $I_{m,sub}^{O,L_O}$ in the range $i \in [0, L_O - L_I]$ and $j \in [0, L_O - L_I]$. At each position (i, j) in $I_{m,sub}^{O,L_O}$, $I_{n,sub}^{I,Li}$ is enlarged by scaling factor $s \in [s_{\min}, s_{\max}]$, and rotated by angle $\theta \in [\theta_{\min}, \theta_{\max}]$ around (i, j) to generated the image $\tilde{I}_{n,sub}^{I,Li}(s,\theta)$. Then $\tilde{I}_{n,sub}^{I,Li}(s,\theta)$ is matched with $I_{m,sub}^{O,L_O}$. The correlation coefficient is adopted as the matching measure because it always ranges from -1 to +1, and is invariant to brightness and contrast. This brightness/contrast invariance can be explained as below [17].

Let $\boldsymbol{x}$ be the column-wise vector obtained by copying the grayscale pixels of $\tilde{I}_{n,sub}^{I,Li}(s,\theta)$, and $\boldsymbol{y}$ be the vector by copying the grayscale pixels in the region of $I_{m,sub}^{O,L_O}$ to be correlated with $\tilde{I}_{n,sub}^{I,Li}(s,\theta)$. Then the brightness/contrast correlation can be written as a least square problem:

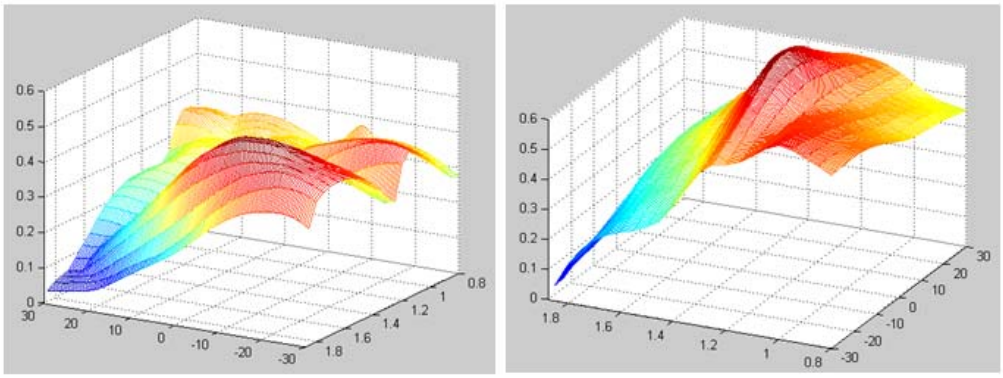
$$\boldsymbol{y} = \beta \boldsymbol{x} + \gamma \boldsymbol{1} + \boldsymbol{\varepsilon} \tag{2}$$

where β and γ is the contrast correction factor and brightness correction factor, respectively, $\boldsymbol{1}$ is a vector of 1's, and $\boldsymbol{\varepsilon}$ is the vector of residual error. The problem is to find β and γ that minimizes $\boldsymbol{\varepsilon}^2$. This problem has a computationally fast solution. Let $\tilde{\boldsymbol{x}} = \boldsymbol{x} - \bar{x}$ and $\tilde{\boldsymbol{y}} = \boldsymbol{y} - \bar{y}$ be the mean-corrected vectors, where $\bar{x}$ and $\bar{y}$ is the mean of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. Then,

$$\beta = \frac{\tilde{\boldsymbol{x}}\tilde{\boldsymbol{y}}}{\tilde{\boldsymbol{x}}^2}, \ \gamma = \bar{y} - \beta \bar{x}, \text{ and } \varepsilon = \tilde{y} - \beta \tilde{x}. \tag{3}$$

The correlation coefficient r_{xy} can be calculated as,

$$r_{xy} = \frac{\tilde{\boldsymbol{x}}\tilde{\boldsymbol{y}}}{\|\tilde{\boldsymbol{x}}\|\|\tilde{\boldsymbol{y}}\|} = \frac{\beta \tilde{x}^2}{\|\tilde{\boldsymbol{x}}\|\|\tilde{\boldsymbol{y}}\|}. \tag{4}$$



(a) O_0^I to O_0^O (b) O_1^I to O_1^O

Fig.4. Object matching measure map for the optical objects and IR object detected in Fig. 3. (a) O_0^I to O_0^O, (b) O_1^I to O_1^O.

This matching for $\tilde{I}_{n,sub}^{I,Li}(s,\theta)$ and $I_{m,sub}^{O,L_O}$ is performed at all position (i, j) for all $s \in [s_{\min}, s_{\max}]$, and $\theta \in [\theta_{\min}, \theta_{\max}]$, where $i \in [0, L_O - L_I]$ and $j \in [0, L_O - L_I]$. At each step, the matching measure shown in Eq. (4) is calculated. The above matching is repeated for all $O_m^O \in O^O$ and $O_n^I \in O^I$, where m = 1, 2, …, M, n = 1, 2, …, N. After this matching process, $M \times N$ measure maps are obtained. And next step is to search these $M \times N$ matching measure maps and find the maximal matching measure peak r_{maxpeak}. For the matching measure map for $O_m^O \in O^O$ and $O_n^I \in O^I$, if the matching measure takes r_{maxpeak} at the scale s_p, rotation angle θ_p, and position (i_p, j_p), then s_p, θ_p, and the translation vector $\boldsymbol{\lambda}_p = (x_m^O - x_n^I + i_p, y_m^O - y_n^I + j_p)^T$ are considered as the best scale, rotation angle, and the translation vector for the matched object pair O_m^O and O_n^I, which is also considered as the best scale, rotation angle, and translation vector for the IR image F_i^I to match the optical image F_i^O, where (x_m^O, y_m^O) and (x_n^I, y_n^I) is the center coordinates of O_m^O and O_n^I, respectively, and $s_p \in [s_{\min}, s_{\max}]$, and $\theta_p \in [\theta_{\min}, \theta_{\max}]$. The matching measure maps for mapping between O_0^I and O_0^O, and O_1^I and O_1^O (others are omitted, here), detected in Fig. 3 (c) and (d), are shown in Fig. 4, where $s \in [0.8, 1.8]$, and $\theta \in [-30^\circ, 30^\circ]$, and at each step s is increased by 0.05, and θ by 0.5. The best matched object pair is O_1^O to O_1^I, and the scale, rotation angle, and translation vector is 1.40, -0.6˚, and $(73, 41)^T$, respectively.

2.2 IMAGE FUSION

After the image registration parameters, s_p, θ_p, and $\boldsymbol{\lambda}_p$ are determined, the image fusion can be performed according to Laplacian pyramid transform [8], the contrast pyramid transform [9], the gradient pyramid transform [10], or DWT [11].

2.2.2 Weighted Image Average Technique

In this technique, first, the scaling, rotation, and translation operations are applied to the IR image F_i^I by employing parameters s_p, θ_p, and $\boldsymbol{\lambda}_p$, to generate the image $\tilde{F}_i^I(s_p, \vartheta_p, \lambda_p)$. Then $\tilde{F}_i^I(s_p, \vartheta_p, \lambda_p)$ and F_i^O are fused according to,

$$F_i^f = \kappa_1 \tilde{F}_i^I(s_p, \vartheta_p, \lambda_p) + \kappa_2 F_i^O \tag{5}$$

where κ_1 and κ_2 are weighting coefficients, and superscript *f* on left hand side means *fuse*. Similarly, DGF response G_i^O and G_i^I of the optical image F_i^O and IR image F_i^I are also fused according to,

$$G_i^f = \kappa_1 \tilde{G}_i^I(s_p, \vartheta_p, \lambda_p) + \kappa_2 G_i^O \tag{6}$$

2.2.3 Discrete Wavelet Transform based Fusion

For an optical image, I_{OP}, and an IR-image, I_{IR}, DWT-based image fusion algorithm can be described by,

$$F = \omega^{-1}(\phi(\omega(I_{OP}), \omega(I_{IR})))$$

where ω is the DWT, ω^{-1} the inverse DWT, ϕ some fusion rules, and F the fused image. That is, I_{OP} and I_{IR} are transformed from normal image space to wavelet coefficients by ω, wavelet coefficients of I_{OP} and I_{IR} are combined by rules ϕ, and the combined wavelet coefficients are transformed to fused image F by ω^{-1}. There is a great variation about the fusion the rules ϕ. The followings are some simple and useful rules.

(i) Take the coefficient with the maximum amplitude from two input wavelet transform arrays;

(ii) Average the values in two input wavelet transform arrays;

(iii) Use the coefficient from image I_{OP} unless the coefficient from image I_{IR} is greater than three times the coefficient from image I_{OP}.

2.3 TARGET DETECTION

So far we use the term *object detection*, but in this section we start naming the same term as *target detection*. They are basically the same, but this report makes a difference in the following sense: object detection means to detect objects from one or two image frames and target detection means to detect objects from a short image sequence. This algorithm employs L frames to localize the target, that is, F_{i-L}^f, F_{i-L+1}^f, …, F_i^f (currently L=10). The target detection algorithm is described as follows.

(i) Detect the specular highlights from the fused DGF response G_k^f (k = i-L, …, i) to locate the objects in the fused image F_k^f, by using the algorithm summarized in Section 2.1.1. The object detected from F_k^f is denoted by $O_{q,k}^f$, where q = 1, 2, …, Q, and Q is the number of objects. $O_{q,k}^f$ is represented by it center coordinates $C_{q,k}^f$, circumscribed rectangle $R_{q,k}^f$, and circumscribed ellipse $E_{q,k}^f$.

(ii) All objects detected from F_{i-L}^f, F_{i-L+1}^f, …, F_{i-1}^f are transformed to image frame F_i^f by using,

$$\tilde{C}_{q,k}^f = \omega_k^{k-1} \times \omega_{k-1}^{k-2} \times \cdots \times \omega_{i-1}^{i} \times C_{q,k}^f \qquad (7)$$

where ω_{m-1}^m is the affine motion from frame m-1 to m, determined at the step of object detection.

(iii) Targets are localized by the grid-clustering method [18]. A filtering operation is applied to the obtained clusters, by thresholding the cluster density with threshold d_{thres}.

Fig.5. Target detection result at frame 23, by using clustering technique to the objects detected from frame 14 to 23.

Fig. 5 shows targets localized by the above clustering algorithm at frame 23, based on the objects detected from frame 14 to 23, where a red dot means an object detected in frame between 14 and 23, green circles mean the clusters obtained, and purple ellipses mean the targets localized.

3. EXPERIMENTAL RESULTS

The algorithm described above were implemented by using Microsoft Visual C++ and Intel Open CV on Windows platforms. The Vivid Datasets provided by the Air Force Research Laboratory were utilized in the experiments performed. The frame interval Δ for object detection is set at 1, the searching range for s and θ in object mapping is set at [0.8, 1.8] and [-30˚, 30˚], respectively, and the increment for s and θ is 0.05 and 0.5˚, accordingly. The weighting coefficient κ_1 and κ_2 for image fusion are both set at 0.5. The image sequence length L for target localization is set at 10. The threshold d_{thres} for cluster filtering is set at 0.65.

Fig. 6 shows some image fusion results using *weighted image average technique*. (a) and (b) shows an optical image and an IR image, respectively, and (c) is the fused image. (d) and (e) is another pair of input images, and (f) is the fused image. From the fused image in (c) and (f), we can see the targets become clear and easy to detect. Especially the pick-up truck in (d) is hidden by the tree shade, but it clearly appears in fused image in (f).

Fig. 7 shows some target detection results. Left column shows the binarization results of the fused DGF responses, and right column the detected targets at frame 10, 57, and 99, respectively. The system outputs the detection results from 10-th frame because the target localization employs 10 frames. The green circles mean the clustering results of the detected objects over 10 frames and the purple ellipses the localized targets. Because the system employs the object detection history (10 frames), the system can still detect the targets although they are lost shortly because the displacement is too small (as shown in (c), the DGF response is zero, *i.e.*, the frame difference is zero).

Fig. 8 shows a DWT-based full image fusion result. (a) is a visible input image, (b) is an IR image, and (c) is full image fusion result. Because the optical image size is 640x480 and an IR image size is 320x256, the images are divided into 16x16 sub-images to conduct DWT and fusion. Fig.9 shows an object-based fusion result. (a) is detected object, and (b) is the fused image. Fig. 10 and Fig. 11 show the image fusion results for another pair of input images.

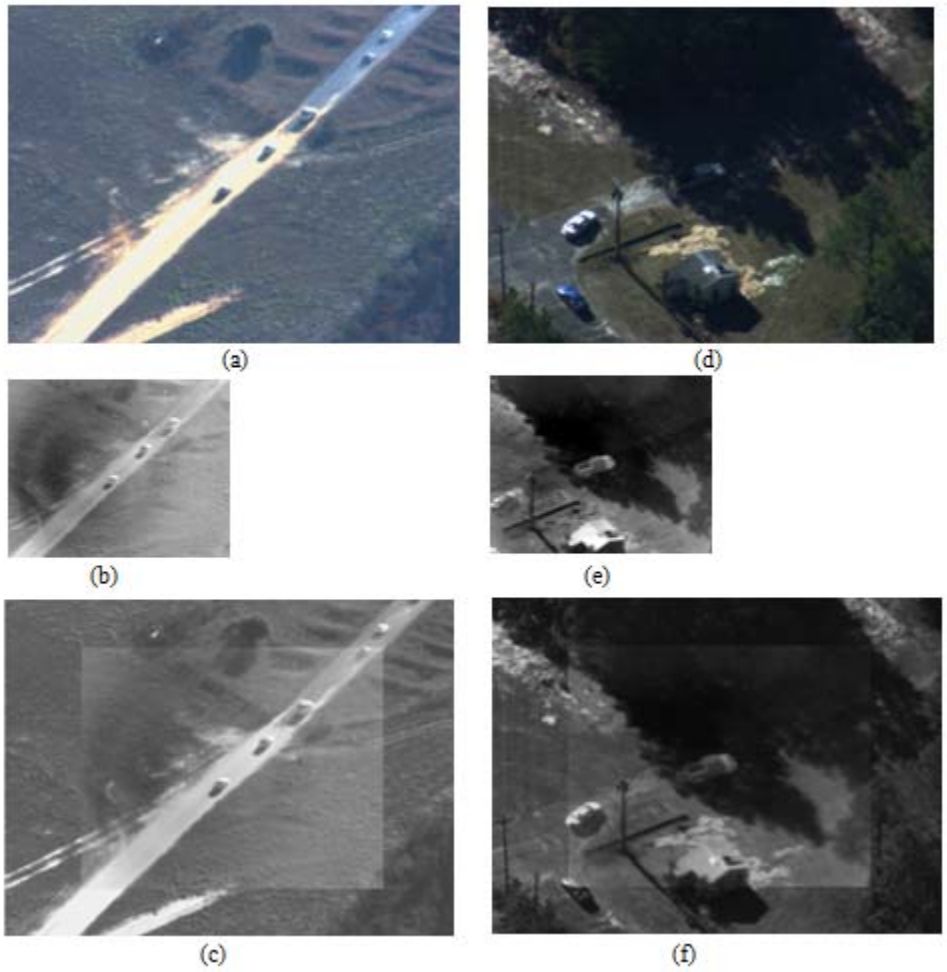


Fig.6. Image fusion results (a) Optical image; (b) IR image; (c) Fusion of image in (a) and (b); (d) Optical image; (e) IR image; (f) Fusion of image in (d) and (e).

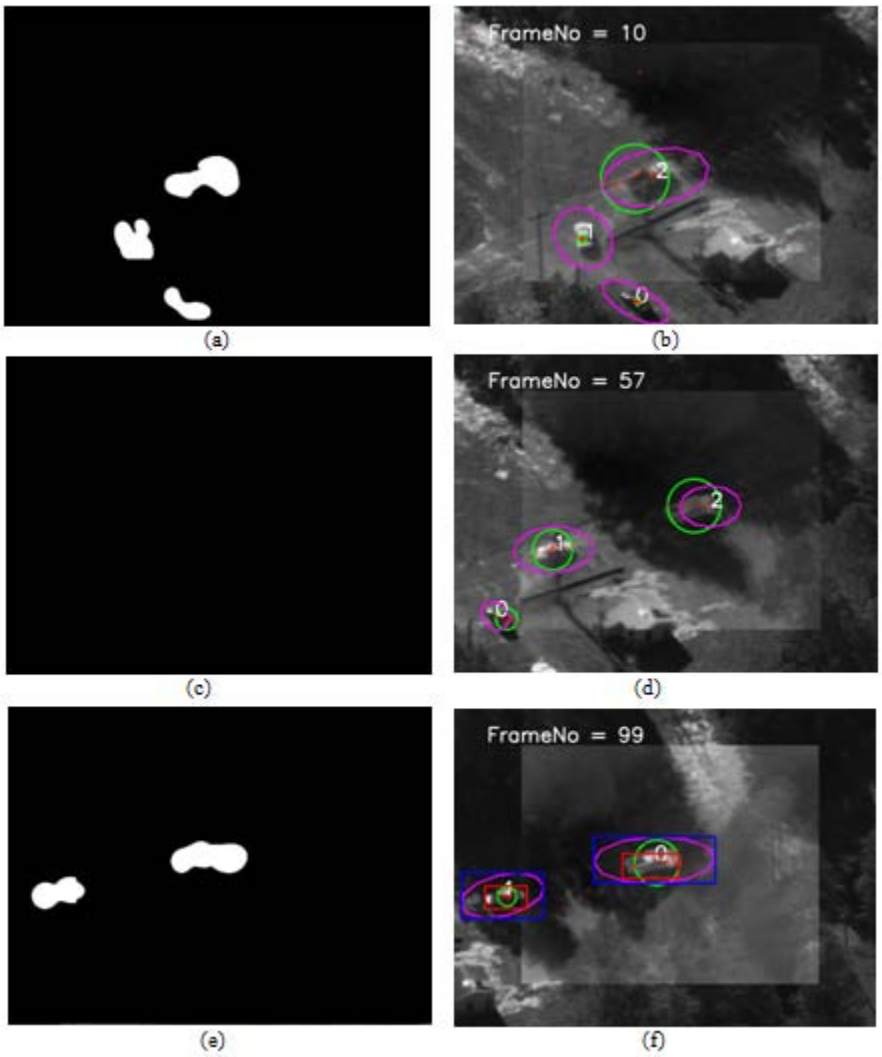
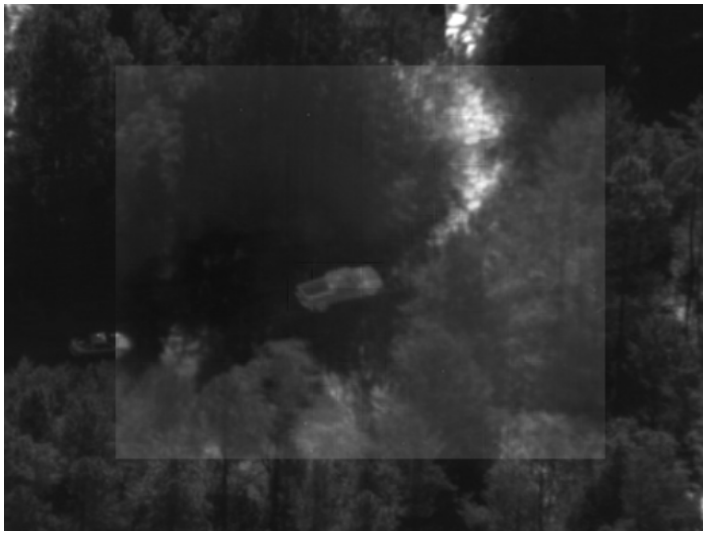


Fig.7. Target detection results in frame 10, 57, and 99. Left column shows the binarization of the fused DGF response, and right column the detected targets.

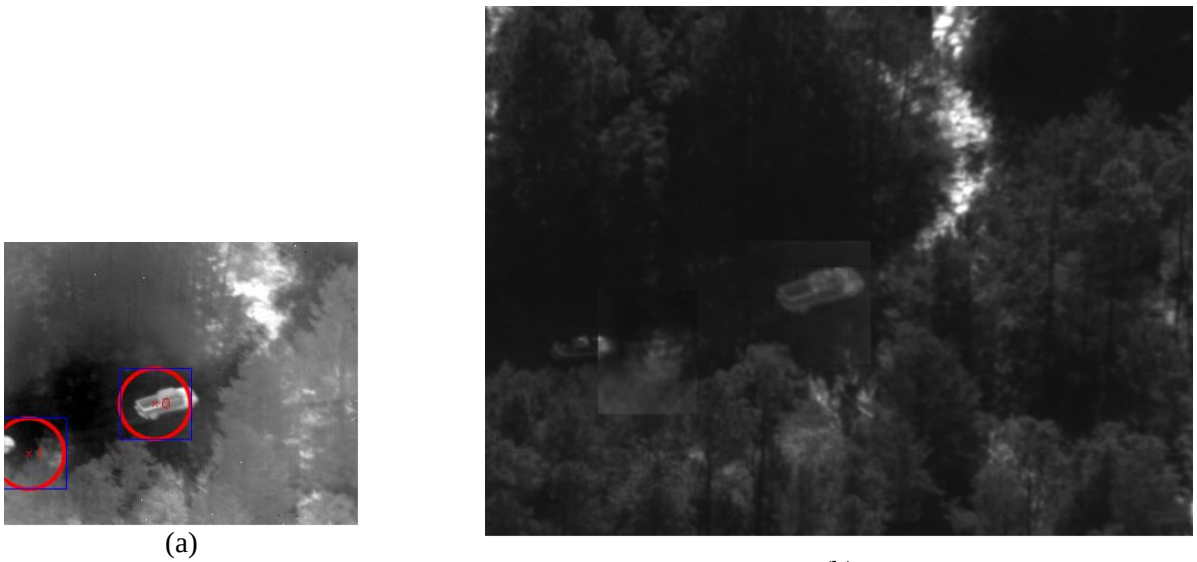
(a)

(b)



(c)

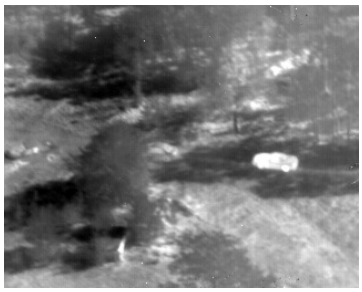
Fig. 8 (a) Visible image, (b) IR image, and (c) fused image.



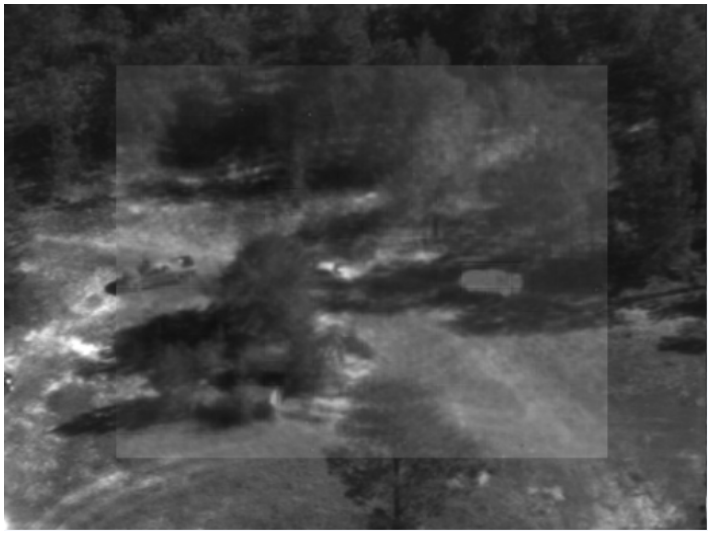
(a) (b)

Fig. 9 (a) Detected objects and (b) object-based image fusion.

(a)

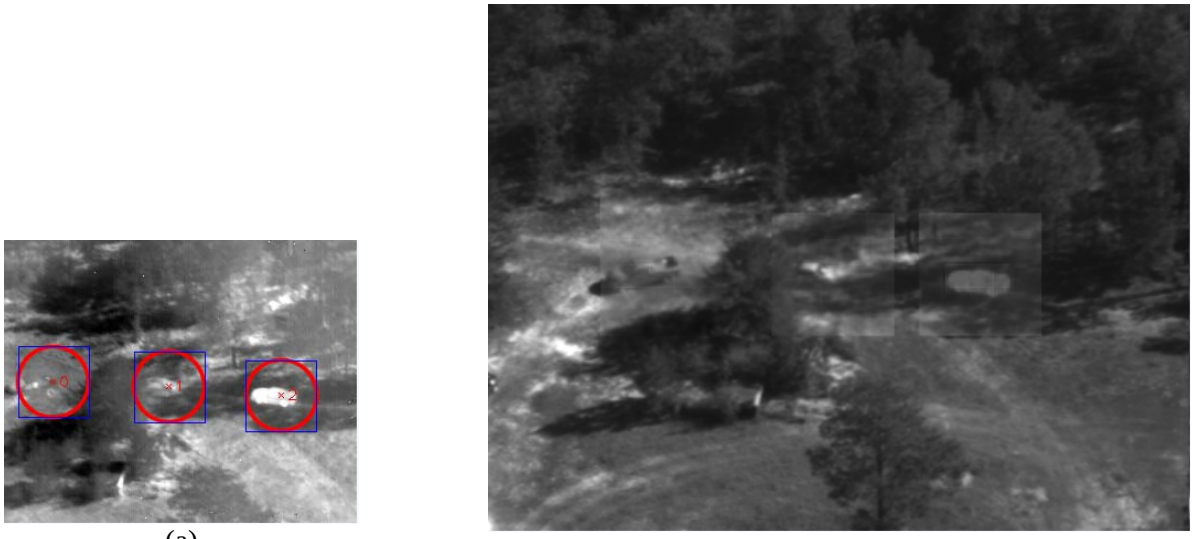


(b)



(c)

Fig. 10 (a) Visible image, (b) IR image, and (c) fused image.



(a) (b)

Fig. 11 (a) Detected objects and (b) object-based image fusion.

4. PERFORMANCE ANALYSIS

The algorithm described above is tested on Windows Vista machine mounted with an Intel Core 2 CPU and 2GB memory, running at 2.33GHz.We employed a 200-frame optical video sequence and a 200-frame IR video sequence to test the performance of the entire algorithm. The videos are sampled at interval $\Delta=2$, *i.e.*, totally it uses 200 optical and IR image frames. The resolution is 640×480 full color for optical image, 320×256 pseudo color for IR image. Each frame contains 2 to 3 objects, totally there are 270 objects. We use the ground truth data to evaluate this algorithm. As shown in Fig. 7 (f), the ground truth target is shown by the red rectangle and the detected target by blue rectangle (the circumscribed rectangle of the detected object). The detected objects are 204. From frame 59 to 70 (totally 36 objects), there are no objects detected because the moving displacement is too small. If we subtract these frames, the total object number becomes 224, the detection rate is 91%. The processing time is shown in Table 1. The image registration/fusion time is 60 seconds, and the average time for target detection is 1.7 seconds per frame.

Table 1. Processing times for image registration/fusion and target detection.

Processing	Specific Task	Time (ms)
Image Registration/ Fusion	Optical object detection	1170
	IR object detection	437
	Object mapping	60403
	Fusion of optical and IR Image	16
	Total	62026
Target Detection (average)	Optical object detection	1171
	IR object detection	426
	Fusion of optical and IR Image	16
	Fusion of DGF response	16
	Target localization	98
	Total	1727

5. CONCLUSIONS AND FUTURE WORK

This report described the development of an (optical and IR) image fusion system for automated target detection and tracking. The system first performs image registration/fusion and then target detection from fused images. Image registration is based on moving object detection and object mapping. Image fusion is based on DWT. The technique for object mapping is invariant to rotation, scale, translation, brightness and contrast. The algorithm for target detection is based on the detection of specular highlights from fused DGF response and clustering technique. The experiment results show this algorithm is valid and efficient. The processing time for image registration/fusion is 60 seconds. This time is acceptable because this processing is executed only once (note that the optical camera and IR camera are mounted on the same moving platform). The average processing time for target detection is 1.7 seconds per frame. This time can be reduced to a half by resizing the optical image to 320×240. Then the performance can be improved to 1.5 frames per second. This speed meets the requirements of many real-time applications.

This research further can be expanded to address the problem of fusion of image sequences collected by independently moving heterogeneous cameras for continuous detection and tracking of moving targets.

6. REFERENCES

1. Hall, D. L., Llinas, J.: An introduction to multisensor data fusion, Proc. IEEE, vol. 85, no. 1, pp. 6-23, Jan. 1997.
2. Varshney, P. K.: Multisensor data fusion, Electronics & Communication Engineering Journal, vol. 9, pp. 245-253, Dec. 1997.
3. Klein, L. A.: Sensor and Data Fusion Concepts and Applications. SPIE, 1993.
4. Ferris Jr., D. D., McMillan, R. W., Currie, N. C., Wicks, M. C., Slamani, M. A.: Sensors for military special operations and law enforcement applications, Proc. SPIE, vol. 3062, pp. 173-180, 1997.
5. Slamani, M. A., Ramac, L., Uner, M., et al: Enhancement and fusion of data for concealed weapons detection, Proc. SPIE, vol. 3068, pp. 8-19, 1997.
6. Franklin, M. R.: Application of an autonomous landing guidance system for civil and military aircraft, Proc. of SPIE, vol. 2463, pp. 146-153, 1995.
7. Kerr, J. R., Pond, D. P., Inman, S.: Infrared-optical multisensor for autonomous landing guidance, Proc. of SPIE, vol. 2463, pp. 38-45, 1995.
8. Burt, P. J., Adelson, E.: The Laplacian pyramid as a compact image code, IEEE Trans. Communications, vol. 31, no. 4, pp. 532-540, Apr. 1983.
9. Toet, A.: Image fusion by a ratio of low-pass pyramid, Pattern Recognition Letters, vol. 9, no. 4, pp. 245-253, 1989.
10. Burt, P. J.: A gradient pyramid basis for pattern-selective image fusion, Society for Information Display, Digest of Technical Papers, pp. 467-470, 1992.
11. Zhang, Z., Blum, R. S.: A categorization and study of multiscale-decomposition based image fusion schemes, Proc. of the IEEE, pp. 1315-1328, Aug. 1999.
12. Sadjadi, F.: Comparative Image Fusion Analysis, Proc. of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05).
13. Sheikh, Y. A., Shah, M.: Trajectory Association across Multiple Airborne Cameras, IEEE Trans. Pattern Anal. Mach. Intell. (accepted).
14. Yao, F. H., Sekmen, A., Malkani, M.: A Novel Method for Real-time Multiple Moving Targets Detection from Moving IR Camera, Proc. of ICPR 2008.
15. Shi, J., Tomasi, C.: Good features to track, Proc. of 9th IEEE Conference on Computer Vision and Pattern Recognition, Springer (1994).
16. Bouguet, J. Y.: Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm, Intel Corporation, 2003.
17. Kim, H. Y., Araújo, S. A.: Grayscale Template-Matching Invariant to Rotation, Scale, Translation, Brightness and Contrast, LNCS, Vol. 4872, pp. 100-113, Dec. 2007.
18. Schikuta, E.: Grid-Clustering: A fast hierarchical clustering method for very large data sets, CRPCTR93358, 1993.